DOCUMENT RESUME

ED 129 893 TM 005 713

AUTHOR van der Kamp, Max; van der Kamp, Leo J. Th.
TITLE Evaluation of Educational Research in the Netherlands.
SPONS AGENCY Netherlands Foundation for Educational Research, The Hague.
PUB DATE Apr 76
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Comparative Analysis; *Educational Research; *Evaluation; Foreign Countries; Measurement Techniques
IDENTIFIERS *Netherlands

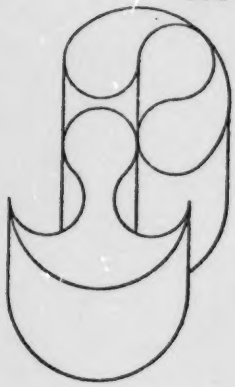
ABSTRACT

To get an impression of the quality of educational research in the Netherlands, 55 paper proposals accepted by the Paper Committee of the Educational Research Day 1974 were evaluated. Each of the 204 judges evaluated two randomly assigned proposals on 27 characteristics. These characteristics were an extension of the instrument used by a committee of the American Educational Research Association in a similar study. The proposals showed a number of specific shortcomings. The general impression was weak. By factor analysis the factorial validity of the instrument was determined. Multiple regression analysis showed the instrument could reasonably predict the general impression of research. (Author/BW)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED129893

12.14



kohnstamm instituut
van de universiteit van amsterdam

EVALUATION OF EDUCATIONAL RESEARCH IN THE NETHERLANDS

Max van der Kamp - Kohnstamm Institute, University of Amsterdam

Leo J. Th. van der Kamp - University of Leiden

Paper presented at the 1976 AERA Annual Meeting
San Francisco, April 19-23, 1976

TM005 713

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

This study was supported by the Netherlands Foundation for Educational Research

keizersgracht 73
postbus 3753
amsterdam-c

EVALUATION OF EDUCATIONAL RESEARCH IN THE NETHERLANDS

Summary

To get an impression of the quality of educational research in the Netherlands 55 paper proposals accepted by the Paper Committee of the Educational Research Day 1974 were evaluated. Each of the 204 judges evaluated 2 randomly assigned proposals on 27 characteristics. These characteristics were an extension of the instrument used by a committee of the AERA in a similar study (Wandt, 1968). The proposals showed a number of specific shortcomings. The general impression was weak. By factor analysis the factorial validity of the instrument was determined. Multiple regression analysis showed the instrument could reasonably predict the general impression of research.

EVALUATION OF EDUCATIONAL RESEARCH IN THE NETHERLANDS

Max van der Kamp - Kohnstamm Institute of Educational Research, University of Amsterdam

Leo J.Th. van der Kamp - Department of Psychology, University of Leiden

Educational research in the Netherlands has been proliferated rapidly in recent years. One indication for this proliferation may give the number of participants in the Annual Meetings of the Dutch Educational Research Association. The first annual meeting was held in 1974 with a total of about 550 participants. In view of the total number of inhabitants of our small country, this amounts to approximately 1 educational researcher to 22000 inhabitants. It should be noticed, however, that the term "educational researcher" is used in this context in a rather loosely defined way. Only recently facilities have been created for a formal training in educational research at our universities. So most of the above mentioned educational researchers can not be considered as a species bred in this new discipline. Another indication for the proliferation of educational research in the Netherlands may be obtained from an examination of the development of the educational research as institutionalized activities. To give you a bird's eye view of the history of post-war educational research, its development will be divided into periods of ten years each.

Period 1: 1950-1960, the early beginnings

In this period educational practice showed its first serious attention for educational research activities as far as the latter products can be used for educational practice. There is also a burgeoning of validation studies with respect to educational instruments. In the second half of the 1950s research and development activities of more than incidental importance have been given thought to more thoroughly than before. Reluctantly funds were made available for educational research.

Period 2: 1960-1970, Dutch educational research growing towards maturity

Particularly in the second half of the 1960s educational research is booming. This sudden increase in educational research activities was coincided with the tremendous increase in the number of enrolled students in our universities. A signal event in this period was the establishment of the Foundation for Educational Research (SVO). This foundation has proven to be an important factor to which the growth of educational research may be ascribed, and it still serves

as a major coördinating force in this research area. In this period, following the developments in the United States, an important study was undertaken on educational opportunity and social equality, the Dutch Project Talent, whereas the first compensatory programs were tried out. At several institutes for tertiary education research teams or special departments were established in order to do research in higher education.

Period 3: 1970- ..., years of stabilization?

Dutch educational research was still in a stage of development in the early 1970s. University graduates, actually the Dutch doctorandi, who had majored in psychology, pedagogy, or sociology and who had just finished their studies took up educational research. These research activities mainly took place at universities, university educational research centres, educational research institutes, or at the newly founded Central Institute for Test Development (CITO), which was modelled after Educational Testing Service. One might say that educational research in the Netherlands has become one of the flourishing "real establishments" nowadays. How educational research shall develop in the years to come, is difficult to foresee. Among others, educational research as well as other fields of research will depend upon the country's economic situation, or rather the economic situation of the developed countries. Education and educational research activities, however, have a low degree of autonomy. It is not a field of political decision-making with its "own" goals and its "own" instruments of policy. To a great extend it depends on the general frame of reference, on the societal goals. Actually, few societal goals are free of educational influence but education is never the only policy instrument for the achievement of such goals. So how educational research in the Netherlands will develop during the second half of this decade also depends upon its impact on educational practice as well as on education at large.

Let us now turn to Dutch educational research and the quality of published educational research in particular. It is a truism to say that an increase of educational information will be of little value for educational practice, unless the quality of research is agreed upon. Of course, the requisite quality and needed form of educational information will vary according to the technical needs and expertise of the target audience (Vockell & Asher, 1974). Our study should be seen as a first attempt to assess the quality of educational research in the Netherlands. Hopefully, it won't give you a too gloomy picture of the present state of Dutch educational research.

OBJECTIVES

The objectives of the investigation were:

(a) to develop an instrument for evaluation of educational research,

(b) to get a general impression of the overall quality of educational research in the Netherlands, and

(c) to identify the specific shortcomings of paper proposals for the 1974 Annual Meeting of the Dutch Educational Research Association as accepted by the Paper Committee.

METHOD

Selection of Material and Judges

Our study was closely related with the 1974 Annual Meeting, where 55 Dutch investigators presented their research findings. The paper proposals of the investigators were used as the material to be evaluated. A random sample of 360 judges was drawn from the total group of participants (mainly educational researchers) at the 1974 Annual Meeting. The judges were mainly psychologists (42%), educationalists (20,6%), sociologists (10,8%) and mathematicans.
Each of the judges had to evaluate two randomly assigned proposals. The evaluation instrument was an extension of that devised by the AERA Committee on Evaluation of Research (Wandt, 1968).
Firstly, each judge was asked to rate the proposals in terms of 27 specific research characteristics. For each characteristic a five-point scale was used, representing five levels of quality: (5) excellent, (4) good, (3) mediocre, (2) poor and (1) complete incompetent.
If the characteristic did not apply to the proposal, the judge was instructed to rate it as 'does not apply'.
Secondly, the judge was asked to rate his general impression of the overall quality of the research reviewed. In addition, the judges were asked whether there was any overlap with other research in the field, to rate his own expertness to evaluate the proposal assigned to him and to rate his acquintance with other publications of the author of the particular paper proposal. The response percentage on the questionaire was 57%. The total number of proposals evaluated and available for analysis was 389.

Data analysis

(1) Frequencies, percentages, means and standard deviations were calculated

to discribe the judges' ratings and to evaluate the quality of the proposals.

(2) Correlation coëfficients were computed for the ratings assigned by the judges to the 27 characteristics, the general impression of the quality of the research reviewed, the subjective expertness of the judges and the reputation of the authors.

(3) Multiple regression analysis was used to examine the interrelations between the 27 characteristics and the general impression of the research reviewed.

(4) The dimensionality of the instrument was identified by principle components analysis followed by VARIMAX rotation.

(5) Differences between the 11 divisions on the 1974 Annual Meeting were examined by multiple discriminant analysis.

Limitations of the Study

The source of the data, the 1974 Annual Meeting of the Dutch Educational Research Association, had some implications for the investigation. An important limitation was the selection of judges and the specific research contributions reviewed, i.e. the specific paper proposals submitted to the 1974 Annual Meeting and accepted by the Paper Committee. These proposals published in the Proceedings of the Annual Meeting were limited in size. This may well be the reason why there was a fair amount of missing data.
The results must be interpreted in the light of the above marks.

RESULTS

The Quality of the Proposals

Means, standard deviations and the total number of observations are presented for each division seperately and for all papers totally (Table 1). The papers were classified into the following divisions:

(A) Analysis of objectives (2)
(B) Assessment procedures (2)
(C) Methodology (4)
(D) Learning and teaching systems (5)
(E) Test development (3)
(F) Evaluation (8)
(G) Innovation (8)
(H) Cognitive functions (8)
(J) Survey research: student characteristics (6)

Table 1. Mean ratings of 28 characteristics of 55 paper proposals all together and for each division seperately. (m=mean, sd=standard deviation and n=number of observations, may vary). For the codes of the divisions, see text, TOT=total.

		A	B	C	D	E	F	G	H	J	K	L	TOT
1. Problem is clearly stated	m :	3.86	3.67	3.26	3.03	3.68	3.68	3.37	3.38	3.07	3.62	3.48	3.43
	sd:	.86	.78	1.23	1.16	.72	.97	1.12	1.06	1.14	.88	1.00	1.05
	n :	14	12	27	33	22	60	54	53	42	39	31	387
2. Hypotheses are clearly stated	m :	3.33	2.78	2.22	2.22	3.11	3.36	2.71	2.44	2.62	2.43	3.29	2.78
	sd:	1.00	.97	1.13	1.22	1.10	1.15	1.31	1.10	1.24	1.03	1.24	1.23
	n :	9	9	23	27	19	53	41	39	29	21	28	298
3. Contribution to theory	m :	3.07	2.78	3.25	2.73	3.00	2.81	2.79	2.94	2.87	2.30	3.00	2.85
	sd:	1.14	.67	.93	1.05	1.08	1.07	1.08	1.11	1.08	.92	.94	1.05
	n :	14	9	28	30	20	59	47	49	39	33	28	356
4. Contribution to educational practice	m :	3.07	4.17	3.00	3.53	3.23	3.45	3.74	3.10	3.24	3.22	2.80	3.32
	sd:	1.21	.72	1.16	.92	1.15	.85	1.11	1.25	1.25	1.16	1.00	1.12
	n :	14	12	28	32	21	60	53	51	42	37	30	380
5. Contribution to societal issues	m :	2.69	3.63	3.04	3.00	2.58	3.23	3.75	2.50	3.59	3.19	3.00	3.16
	sd:	1.32	1.21	1.13	.91	.96	.99	1.15	1.09	1.18	1.12	1.26	1.17
	n :	13	11	27	30	19	56	52	46	41	36	30	361
6. Assumptions are clearly stated	m :	3.08	3.56	2.74	2.32	2.79	2.57	2.25	2.47	2.22	2.66	2.74	2.52
	sd:	1.08	.88	1.06	1.14	.71	1.14	1.23	1.14	1.12	1.06	1.15	1.12
	n :	12	9	27	31	19	56	48	49	36	35	31	353
7. Limitations of the study are stated	m :	2.79	2.18	2.23	2.03	2.28	2.81	1.92	1.75	2.56	2.78	2.37	2.33
	sd:	.70	.60	1.14	.85	.90	1.08	.94	.94	.84	1.10	1.07	1.04
	n :	14	11	26	30	18	58	48	47	41	36	30	359
8. Important terms are defined	m :	2.69	2.58	2.47	2.81	2.74	2.58	2.42	2.28	2.63	2.74	2.24	2.54
	sd:	.85	.79	1.20	1.00	.81	1.09	1.11	1.07	1.09	.93	.95	1.04
	n :	13	12	28	32	19	58	48	50	41	34	29	364
9. Relationship of the problem to previous research is made clear	m :	2.85	1.78	2.33	1.80	2.44	2.13	2.15	2.40	2.46	2.49	2.46	2.29
	sd:	.90	1.09	1.11	.85	1.04	1.16	1.06	1.19	1.21	1.12	1.29	1.13
	n :	13	9	27	30	18	55	47	47	39	33	28	346

Table 1, continued

10. Research design is described fully	m :	3.29	3.00	2.52	2.52	2.73	3.28	3.18	3.19	3.15	3.21	3.16	3.07
	sd:	.91	.85	1.16	1.00	.94	.96	1.14	1.12	.96	.92	1.10	1.05
	n :	14	12	25	31	22	60	50	52	41	39	31	377
11. Research design is appropriate to the solution of the problem	m :	3.21	3.55	3.10	2.85	3.17	3.16	3.02	2.83	3.08	3.24	3.00	3.07
	sd:	.89	.52	.83	.86	.79	.97	1.13	1.15	1.06	.93	.93	.99
	n :	14	11	21	27	18	57	44	47	37	37	29	342
12. Research design is free of specific weaknesses	m :	3.07	2.82	2.91	2.60	2.88	3.20	2.80	2.89	2.97	3.03	3.03	2.95
	sd:	.92	.75	.92	.96	.96	.89	1.03	1.05	.77	.97	.93	.94
	n :	14	11	22	25	16	59	50	47	36	36	30	346
13. Population and sample are described	m :	2.83	2.67	2.10	2.65	2.25	2.60	2.76	2.31	3.31	3.24	2.37	2.70
	sd:	.72	.89	1.22	1.19	1.49	1.03	1.14	1.06	1.05	.94	1.13	1.12
	n :	12	12	21	23	8	60	46	45	42	38	30	337
14. Method of sampling is appropriate	m :	2.38	2.38	1.82	2.65	2.50	2.37	2.15	1.61	3.00	2.78	1.72	2.32
	sd:	.92	.92	.98	1.41	1.31	.99	.96	.76	1.17	1.18	.90	1.12
	n :	8	8	11	17	8	41	34	31	30	32	18	238
15. Data-gathering methods or procedures are appropriate to the solution of the problem	m :	3.00	3.08	3.13	2.96	2.86	3.10	2.78	2.76	3.11	3.03	3.03	2.97
	sd:	1.00	.90	.89	1.10	1.03	.87	.94	1.08	.88	.97	.91	.95
	n :	11	12	16	25	14	52	46	42	37	36	29	320
16. Data-gathering methods or procedures are utilized correctly	m :	3.40	3.08	2.81	3.09	2.71	3.07	2.83	3.11	3.22	3.30	3.00	3.06
	sd:	.70	.67	.91	1.08	.99	.88	1.00	.90	.98	.64	.78	.89
	n :	10	12	16	23	14	46	36	35	32	33	24	281
17. Reliability of the procedures used	m :	3.36	2.80	3.06	2.92	2.64	2.71	2.57	2.72	3.04	2.76	2.92	2.81
	sd:	1.03	.92	.83	1.02	1.01	.94	1.01	1.00	.85	.95	.86	.95
	n :	11	10	17	26	14	51	35	36	27	29	25	281
18. Validity of the evidence gathered	m :	3.00	2.82	2.81	2.86	2.53	2.29	2.69	2.52	3.21	2.79	2.74	2.70
	sd:	.89	1.08	.83	1.03	.74	.80	1.00	.97	.94	1.00	.81	.95
	n :	11	11	16	29	15	48	39	33	29	28	23	282
19. Appropriate methods are selected to analyze the data	m :	3.91	2.75	3.20	2.68	3.00	3.10	2.89	3.28	3.50	3.21	3.25	3.15
	sd:	1.00	.87	.95	1.13	.89	.97	1.10	.91	1.07	.81	.70	.98
	n :	12	12	20	22	11	50	37	36	28	34	28	290
20. Methods utilized in analyzing the data are applied correctly	m :	4.00	2.80	3.00	2.50	2.75	3.21	2.70	3.11	3.48	3.11	3.29	3.08
	sd:	.85	.79	1.11	1.01	1.14	1.04	1.21	1.03	1.16	.83	.75	1.06
	n :	12	10	19	22	12	47	23	28	33	28	24	258

Table 1, continued

21. Results of the analysis are presented clearly	m :	3.00	3.09	2.43	2.41	2.65	2.95	2.44	2.31	2.90	2.56	2.80	2.68
	sd:	1.04	1.14	1.03	1.05	.86	1.09	1.18	.99	.96	1.05	1.01	1.06
	n :	14	11	21	29	17	59	36	35	40	34	31	327
22. Conclusions are clearly stated	m :	3.21	3.33	2.70	3.00	3.21	3.19	2.22	2.74	3.10	2.84	2.48	2.89
	sd:	1.19	1.23	.93	.98	.81	1.04	1.18	1.08	1.10	1.08	1.06	1.10
	n :	14	12	23	26	22	59	32	38	39	32	31	328
23. Conclusions are substantiated by the evidence presented	m :	3.29	2.73	2.60	2.66	2.56	3.16	2.13	2.30	2.74	2.63	2.86	2.71
	sd:	.91	1.01	1.05	1.06	.96	1.02	1.04	.95	1.09	1.15	.89	1.06
	n :	14	11	20	26	16	56	30	30	31	27	28	289
24. Generalizations are confined to the population from which the sample was drawn	m :	2.67	2.82	3.24	2.69	2.72	2.30	2.63	2.08	3.12	2.64	2.24	2.58
	sd:	1.07	1.17	.66	1.06	1.07	.94	1.00	.93	.95	1.14	.88	1.03
	n :	12	11	17	32	18	57	46	39	34	33	25	324
25. Report is clearly written	m :	3.23	2.82	2.54	2.73	2.56	2.95	2.83	2.55	2.85	2.90	2.77	2.79
	sd:	1.17	.98	.95	1.21	.98	1.02	1.18	1.02	1.09	1.03	1.06	1.07
	n :	13	11	26	33	18	60	46	47	39	38	31	362
26. Report is logically organized	m :	3.71	3.17	2.93	3.03	3.26	3.53	3.20	2.85	3.15	3.50	3.45	3.24
	sd:	.73	1.12	.96	1.12	.81	.92	1.01	1.00	1.07	.92	1.03	1.00
	n :	14	12	27	32	19	59	50	47	39	38	31	368
27. Tone of the report displays an unbiased, impartial scientific attitude	m :	3.57	3.40	3.42	3.45	3.56	3.44	3.30	3.36	3.39	3.42	3.53	3.42
	sd:	.76	.70	.70	.74	.71	.81	.89	.65	.77	.77	.57	.74
	n :	14	10	26	29	18	55	43	45	36	36	30	342
28. General impression of this research	m :	3.36	3.25	3.00	2.78	2.85	3.10	2.96	2.80	3.13	3.10	3.03	3.01
	sd:	1.08	.87	.83	.98	.67	.95	1.03	1.12	1.07	.97	.96	.98
	n :	14	12	27	32	20	60	52	45	40	39	30	371

Shortcomings of the paper proposals

From the last column of Table 1 it may be concluded that the judges have been rather lenient with respect to their judgement of the paper proposals. In general the formulation of the problem was clear enough, while the way of reporting the investigation was unbiased as well. Furthermore the total stock of studies was judged to be more pertinent to educational practice than to theory construction. On 16 out of the 28 characteristics used, the total number of papers were rated as insufficient. Among others, the limitations of the study were not clearly stated, neither was the relationship with previous research made clear nor the assumptions. A clear definition of the most important terms used by the investigators was lacking. Other shortcomings had to do with sampling techniques, the validity of the information gathered, the presentation of the analysis, the way the conclusions were drawn from the material gathered and the generalizibility of the research findings. The general impression of the total stock of papers presented at the 1974 Annual Meeting of the Dutch Educational Research Association, was rather weak.

A Comparison with the Results of the American Study

In order to facilitate a crude comparison the ratings of the Dutch research proposals were compared with the ratings published by Wandt (1968). The latter ratings were based on a similar American study where 81 articles on educational research were judged with respect to their quality.

In Table 2 the mean ratings of the American articles on educational research are juxtaposed to those of the Dutch paper proposals. From Table 2 it can be seen that, apart from some differences, the similarities in the findings are remarkable. The differences have to do with the formulation of the hypotheses, the description of the population studied and with the sampling procedure. Although the specific nature and subject matter of the American articles on educational research are unknown to us, it might well be that much hypothesis-testing research is included, and consequently more attention is paid to the formulation of the hypotheses. Another difference is the way of reporting and presenting the findings and the conclusions. Neither is this difference very remarkable, bearing in mind that the American study used journal articles, while paper proposals were used in the Dutch study. It is a well known fact that that manuscripts for publication are heavily screened; paper proposals, on the other hand, are generally more loosely formulated as they pertain to

Table 2 Mean ratings of 81 American articles on educational research and of 55 paper proposals for the 1974 Annual Meeting of the Dutch Educational Research Association.

characteristics	mean rating	
	81 American articles on Educational Research	55 paper proposals 1974 Annual Meeting DER
1. Problem is clearly stated	3,41	3,43
2. Hypotheses are clearly stated	3,04	2,78
3. Contribution to theory		2,85
4. Contribution to educational practice	3,31	3,32
5. Contribution to societal issues		3,16
6. Assumptions are clearly stated	2,40	2,52
7. Limitations of the study are stated	2,41	2,33
8. Important terms are defined	2,84	2,54
9. Relationship of the problem to previous research is made clear	2,60	2,29
10. Research design is described fully	3,03	3,07
11. Research design is appropriate to the solution of the problem	2,65	3.07
12. Research design is free of specific weakness	2,42	2,95
13 Population and sample are described	3,18	2,70
14. Method of sampling is appropriate	2,85	2,32
15. Data-gathering methods or procedures are appropriate to the solution of the problem	2,99	2,96

Table 2, continued

16. Data-gathering methods or procedures are utilized correctly	3.01	3,06
17. Reliability of the procedures used	2,49	2,81
18. Validity of the evidence gathered		2,70
19. Appropriate methods are selected to analyse the data	2,83	3,15
20. Methods utilized in analyzing the data are applied correctly	3,11	3,08
21. Results of the analysis are presented clearly	3,11	2,68
22. Conclusions are clearly stated	3,06	2,89
23. Conclusions are substantiated by the evidence presented	2,63	2,71
24. Generalizations are confined to the population from which the sample was drawn	3,07	2,58
25. Report is clearly written	3,21	2,79
26. Report is logically organized	3,46	3,24
27. Tone of the report displays an unbiased, impartial scientific attitude	2,42	3,42

research studies the conclusions of which are not final.
The American and the Dutch experts who evaluated the papers do agree with respect to the formulation of the problem, the significance and the objectivity of the research report. In the Dutch study as well as in the American one the formulation of the limitations of the investigation, the definition of the most important terms, and the relations with earlier research turned out to be weak.
Finally, the findings in our study are similar to those of the Wandt study in the sense that the way conclusions were drawn from the material gathered, was judged to be unacceptable.

Authors' reputation

In our study the judges were asked to rate on a five-point scale whether they were familiar with other publications of the authors of the paper proposals. The mean of these ratings was 1.48 (with extreme upper-value 2.56). This finding might imply at least that either the authors of the paper proposals for the Annual Meeting have a low publication rate, or that the judges are not familiar enough with and neither keep themselves informed of what is going on in the field of educational research in the Netherlands. We are of the opinion that the latter explanation seems more plausible than the former, the more so in view of the tentative conclusions of the Examining Committee for Social Research. The latter committee came to the conclusion that much social research is badly documented.

Subjectivly Perceived Expertness

The paper proposals were randomly assigned to the judges. The question however, is whether the judges consider themselves to be sufficiently qualified to assess the paper proposals' quality. Therefore the judges were asked to rate their own expertness as to to the evaluation of the assigned paper proposals. Again ratings were given on a five-point scale. The mean of the subjectively perceived expertness was 2.72 with a standard deviation s=0.87. So the judges in general perceived themselves as not being qualified enough to evaluate the paper proposals. One may ask whether educational research has developed into such diverse areas, that no one dares to call himself an expert in whatever specific field. Possibly a number of the non-respondents did not cooperate for this very reason. To investigate whether the same holds if the overall ratings are split up into the seperate divisions of educational research the mean of the ratings are calculated for each of the eleven divisions. (Table 3).

Table 3. The subjectively perceived expertness of the judges for each division

division	mean	standard dev.
Survey research: teacher characteristics	3.00	0.67
Analysis of objectives	2.86	1.10
Innovation	2.85	0.72
Evaluation	2.82	0.83
Learning and teaching systems	2.82	0.88
Assessment procedures	2.75	1.06
Cognitive functions	2.63	0.82
Miscellaneous	2.61	1.02
Survey research: student characteristics	2.58	0.92
Test development	2.50	1.01
Methodology	2.43	0.84

Overlap with other Research

To get an idea whether there was much overlap in the research, the judges were asked whether overlap did actually occur. 24.6% of the judges affirmed that there is overlap in research. Not all overlap in research, however, is inefficient -replications of investigations might be very useful- 24.6% overlap is perhaps an alarming percentage. This overall overlap in research can also be split up into the above mentioned 11 divisions. The overlap percentages for the separate divisions appear in Table 4.

Agreement between Judges

Agreement between judges with respect to a single paper proposal differed considerably. The mean intercorrelation varied between r=0.05 and r=0.54. Total interrater agreement turned out to be r=0.26, which is a poor overall interjudge correlation. Such a low interrater agreement should not surprise us. Goldberg (1968) e.g., in research on clinical judgement, found a median correlation of r=0.38 between experts who had to judge the severity of ulcers. Expertness was no guarantee for consensus according to Goldberg. To study

Table 4. Percentage of judges indicating overlap in research.

divisions	percent
Learning and teaching systems	40.0
Assessment procedures	31.5
Survey research: student characteristics	30.8
Evaluation	29.6
Innovation	27.6
Analysis of objectives	25.4
Test development	22.5
Miscellaneous	21.6
Survey research: teacher characteristics	18.1
Cognitive functions	14.3
Methodology	11.3

whether the subjectively perceived expertness is independent of general agreement between the judges, the rank correlation was calculated between subjectively perceived expertness on the one hand, and the mean interrater agreement for a single paper proposal on the other. This yielded a correlation coëfficient r_s=0.09. Neither in our study is expertness of judges (at least subjectively perceived expertness) related with consensus. Apparently the norms used to evaluate research still differ a great deal among "experts".

Discriminant Analysis between Divisions

Statistical testing of the means and standard deviations of the divisions on each of the characteristics separately did not yield any significant difference. The difference between the 11 divisions can also be computed using the analysis of discriminance. Discriminant analysis yielded 4 discriminant functions which accounted for 73.6% of the total variance. Only the first and the second discriminant functions were statistically significant, together explaining 50,3% of the total variance. The latter two discriminant functions gave two new so called discriminant variables which may tentatively be interpreted as "scope versus

precision" and "theoretical orientation versus practical orientation". Figure 1 shows both discriminant functions with a plot of the 11 divisions in this two-space. Clearly the 11 divisions differ considerably among each other.

Figure 1 Insert about here

Dimensionality of the Instrument

To study whether the instrument used to evaluate the paper proposals tapped different fundamental aspects, the matrix of correlations between the 28 characteristics was factor analyzed (principal component analysis). In table 5 the factor loadings of the 28 characteristics (only factor loadings above 0.40 are reported) on 5 factors rotated according the VARIMAX criterion are given. These 5 factors accounting for 62% of the total variance, may be interpreted as follows.

I Results/conclusions/reporting (variance explained 18%)
II Research methodology (variance explained 17%)
III Formulation of the problem in a wider scope (variance explained 11%)
IV Significance for education (variance explained 7%)
V Description of the population/sampling (variance explained 7%)

These findings are in agreement with the results obtained from a hierarchical cluster analysis of the same data.

Insert tabel 5 about here

The validity of the Instrument

Multiple-regression analysis was used to study whether the 'general impression of the research' (characteristic 28) could be predicted from the other 27 characteristics. This multiple correlation turned out to be R=0.83, and the corresponding variance explained 69%.

And a tentative interpretation of the multiple-regression analysis might be that highly qualified educational research is, according to the judges in our study, relevant research of practical significance, methodologically well-designed and clearly presented on the part of the researcher. How far the 27 characteristics can be used actually to get an general impression of a research report may not be answered as yet. A cross validation needs to be undertaken to answer this question.

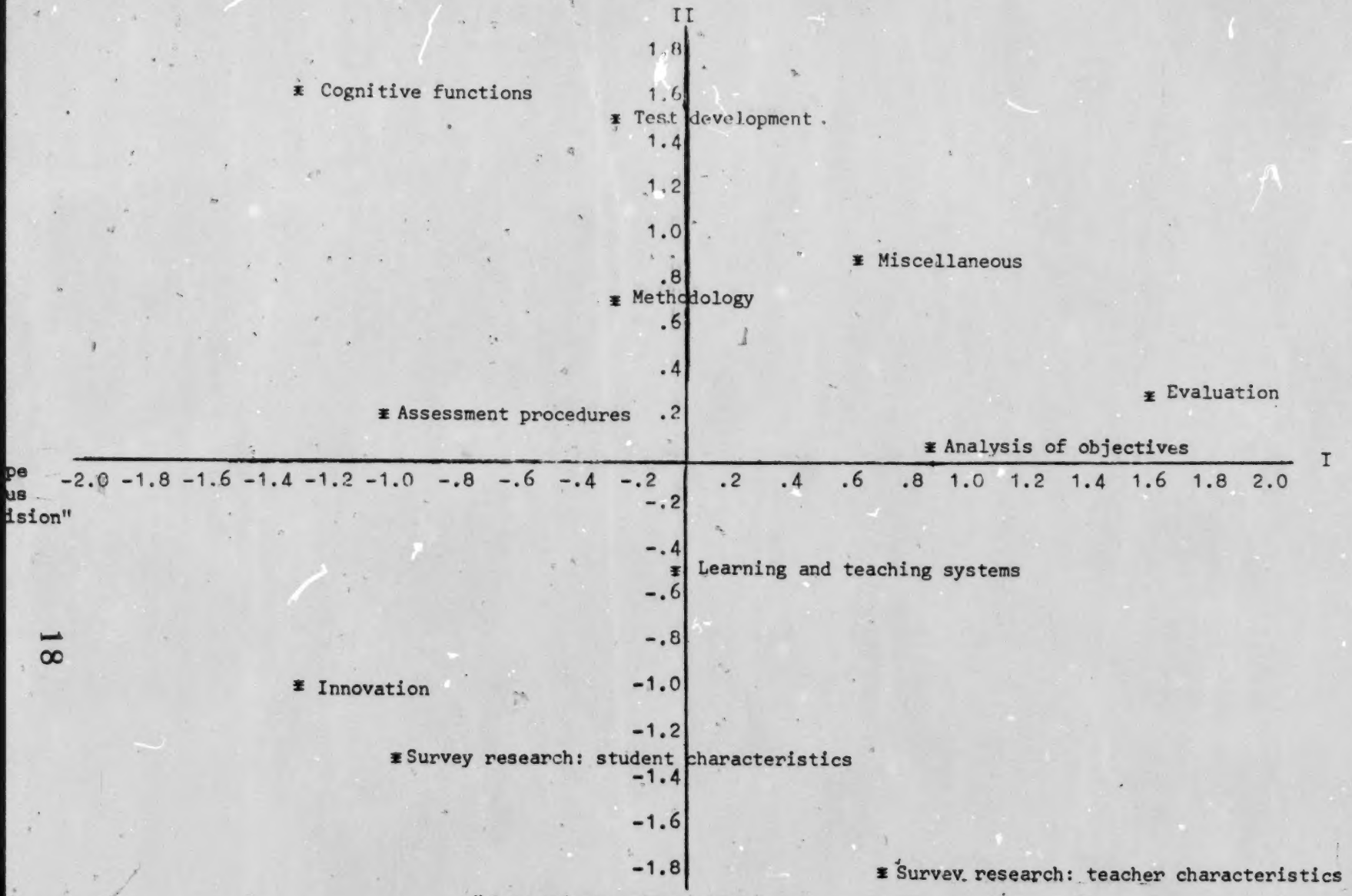


Figure 1. 11 Divisions of the 1974 Annual Meeting in a two-dimensional space

Table 5. Factor loadings of rotated factor solutions. Only loadings above 0.40 are included (decimal points are omitted)

Characteristics	Factors I	II	III	IV	V
1. Problem is clearly stated	53				53
2. Hypotheses are clearly stated	66				43
3. Contribution to theory	45		54		
4. Contribution to educational practice			84		
5. Contribution to societal issues			78		
6. Assumptions are clearly stated	75				
7. Limitations of the study are stated				41	40
8. Important terms are defined	54				
9. Relationship of the problem to previous research is made clear	63				
10. Research design is described fully					58
11. Research design is appropriate to the solution of the problem		44			
12. Research design is free of specific weakness		57			41
13. Population and sample are described				79	
14. Method of sampling is appropriate				82	
15. Data-gathering methods or procedures are approporiate to the solution of the problem.		70			
16. Data-gathering methods or procedures are utilized correctly		74			
17. Reliability of the procedures used		76			
18. Validity of the evidence gathered		68			
19. Appropriate methods are selected to analyze the data		80			
20. Methods utilized in analyzing the data are applied correctly		78			
21. Results of the analysis are presented clearly					69
22. Conclusions are clearly stated					73
23. Conclusions are substantiated by the evidence presented.		41			67
24. Generalizations are confined to the population from which the sample was drawn			50		

Table 5, continued	Factors I	I	III	IV	V
25. Report is clearly written					71
26. Report is logically organized					75
27. Tone of report displays an unbiased, impartial, scientific attitude					49
28. General impression of this research	(26)	(39)	(35)	(10)	59
Variance explained	11%	17%	9%	7%	18%

DISCUSSION

The Perception of Educational Research in the Netherlands

We have outlined the shortcomings of our study earlier. The material judged consisted of a cross section of educational research presented at a certain time, while the group of judges was choosen in a haphazard way. Therefore it does not seem justified to generalize our findings in order to make statements with respect to educational research in the Netherlands at large.
Nevertheless the results of our study are both distressing and alarming. The more so as the Dutch educational system is changing. Educational research might contribute to a more rational change of the present educational system. Policy decisions with respect to the change of the educational system should more rely on sound and methodologically well designed research that it has been until now.
The findings of our study suggest the following recommendations:

(a) the consumers of educational research information should adopt a more critical attitude towards the quality of educational research;
(b) educational research reports should be thoroughly evaluated, and
(c) one should take care for a too rapid and careless dissemination of educational information.

These recommendations are based on the tacit assumption that educational decision makers as well as teachers will use the information obtained in educational research in formulating their policy guidelines as well as in their teaching activities.

The Evaluation Instrument

The instrument used seems to be useful for a critical evaluation of educational research papers. As can been seen from the results of the factor-analytic procedure, groups of items of the evaluation form may be distinguished with each group tapping a specific aspect of research papers. These are the following specific aspects of the paper proposals' quality,

I. Results, conclusions and clarity of reporting
II. Adequate design
III. Clarity of the problem, and of the hypotheses
IV. Significance for education
V. Description of the population and representativeness of the sample.

The items of the form also discriminate between the 11 divisions, hence some discriminant validity might be claimed for this evaluation form for research papers. And it turned out also that the perceived overall quality of a paper proposal can be predicted rather accurately from the individual items.

REFERENCES

Contouren van een toekomstig onderwijsbestel

Discussie nota, Tweede Kamer, zitting 1974-1975, 13459, nrs. 1-2

Council of Europe, Educational research policy in European Countries. Documentation Centre for Education in Europe, 1973, Survey Strassbourg, 1974

Council of Europe. Report on the training and career structures of Educational Researchers. Strassbourg. Documentation Centre for Education in Europe, 1974

De Groot, A.D.
Some remarks on educational measurement and research in the Netherlands,1974-1975. RITP-memorandum, 047. Amsterdam, RITP, 1974

Goldberg,L.R.
Simple models or simple processes? Some research on clinical judgement. American Psychologist, 1968, 23, 483-496

Maini, S.M. and B. Nordbeck.
Critical moments, the creative process and research motivation. A discussion of results based on empirical studies of scientists. International Social Science Journal, 1973, (25), 190-204

Runkel, F. and J. Mc. Grath.
Research on Human Behaviour: A systematic Guide to Method New York, Holt, 1972

Shay, C.B., W.S. Zimmerman and W.B. Michael.
The factorial validity of a rating scale for the evaluation of research articles. Educational and Psychological Measurement, 1972, 32, 453-457

Stanley, J.C.
Improving Experimental Design and Statistical Analysis. Chicago, Rand McNally, 1967

Stichting voor Onderzoek van het Onderwijs, Onderzoek en Onderwijsbeleid Verslag Studiedag Onderwijsresearch. Den Haag 17 oktober 1969. Groningen, Wolters Noordhoff, 1970

Strauss,S.
Guidelines for Analysis of Research Reports. Journal of Educational Research 1969, (63), 165-169

Verkenningscommissie Sociaal Onderzoek, Sociaal-Wetenschappelijk onderzoek en beleid, Nijmegen, 1975

Van der Kamp, M.
Evaluatie Onderwijsresearch, Amsterdam, Kohnstamm Instituut 1975

Walstijn, W. van, L.J.Th. van der Kamp en M. van der Kamp
Kansen op Onderwijs. een literatuurstudie over ongelijkheid in het Nederlandse onderwijs. Den Haag, Voorlopige Wetenschappelijke Raad voor het Regeringsbeleid, 1975.

Vockel, E.L. and W. Asher
Perceptions of Document Quality and Use by Educational Decision Makers and Researchers. American Educational Research Jcurnal, 1974, 11, 249-258

Wandt, E.
ACross-Section of Educational Research.New York, David McKay Company, 1968

Wardt A.W., B.W. Hall and C.F. Schramm
Evaluation of Published Educational Research: A national Survey. American Educational Research Journal, 1975, 12, 109-128

Werkgroep Onderwijsresearch. Onderwijsresearchdag 1974. Groningen, Tjeenk Willink, 1975

Authors

Max van der Kamp (born 2-5-'47) studied psychology at the University of Amsterdam. Works since 1971 at the Kohnstamm Institute of Educational Research, University of Amsterdam.
Publications on curriculumresearch and educational evaluation, participated in a study about inequality in the Dutch educational system. Is working on a doctoral dissertation about aims and objectives of arts in secundary education.

Leo J.Th. van der Kamp (born 19-12-37) studied psychology. Degree Ph.D. University of Leiden. His present position is reader in research methodology at the University of Leiden. Specialization educational and psychological research methodology, test theory, multivariate analysis.